OPEN ACCESS Freely available online

J. Johansson, A. Mäkivirta, M. Malinen and V. Saari, "Interaural Time Difference Prediction Using Anthropometric Interaural Distance"
J. Audio Eng. Soc., vol. 70, no. 10, pp. 843–857, (2022 October).
DOI: https://doi.org/10.17743/jaes.2022.0038

Interaural Time Difference Prediction Using Anthropometric Interaural Distance

JAAN JOHANSSON,[2] **AKI MÄKIVIRTA,**[1,*] *AES Fellow*, **MATTI MALINEN,**[2] **AND**
(jaan.johansson@kuava.fi) (aki.makivirta@genelec.com) (matti.malinen@kuava.fi)

VILLE SAARI,[1] *AES Associate Member*
(ville.saari@genelec.com)

[1]*Genelec Oy, Iisalmi, Finland*
[2]*Kuava Oy, Kuopio, Finland*

This paper studies the feasibility of predicting the interaural time difference (ITD) in azimuth and elevation once the personal anthropometric interaural distance is known, proposing an enhancement for spherical head ITD models to increase their accuracy. The method and enhancement are developed using data in a Head-Related Impulse Response (HRIR) data set comprising photogrammetrically obtained personal 3D geometries for 170 persons and then evaluated using three acoustically measured HRIR data sets containing 119 persons in total. The directions include 360° in azimuth and –15° to 60° in elevation. The prediction error for each data set is described, the proportion of persons under a given error in all studied directions is shown, and the directions in which large errors occur are analyzed. The enhanced spherical head model can predict the ITD such that the first and 99th percentile levels of the ITD prediction error for all persons and in all directions remains below 122 μs. The anthropometric interaural distance could potentially be measured directly on a person, enabling personalized ITD without measuring the HRIR. The enhanced model can personalize ITD in binaural rendering for headphone reproduction in games and immersive audio applications.

0 INTRODUCTION

The Head-Related Transfer Function (HRTF) and its time-domain equivalent Head-Related Impulse Response (HRIR) characterize the sound localization of a person. Sound localization with humans uses both binaural and monaural acoustical cues, both contained in the HRTFs [1–3].

Interaural time difference (ITD) is a function of space, changing with sound direction in azimuth and elevation. ITD models are generalizations of the ITD for a given head size describing ITD as a function of the angle of sound arrival. It is a great advantage in applications requiring binaural rendering if personal ITD can be predicted with sufficient accuracy without having to know the person's full HRTF, such as in gaming or when access to completely personal HRTF data is not available.

Woodworth and Schlosberg [4] derived an ITD model for a rigid sphere of radius r, or the spherical head (SH) model. The ITD is based on the difference between propagation path lengths of the sound to each ear. The zero elevation ITD as a function of azimuth angle ϕ expressed in degrees for a head radius r is

$$\tau_W(\phi) = \begin{cases} \frac{1}{c} r(\phi + \sin\phi) & 0 \le \phi \le 90 \\ \frac{1}{c} r(180 - \phi + \sin\phi) & 90 \le \phi \le 180, \end{cases} \quad (1)$$

where c is the speed of sound in air.

Kuhn [5] continued with the SH model and derived theoretical ITD at zero elevation from the analytical solution for scattered wave at the surface of a rigid sphere. He also obtained different ITD values for low (<500 Hz) and high (>2 kHz) frequency sound. The Kuhn ITD model is

$$\tau_K(\phi) = \frac{3r\sin\phi}{c}, \quad < 0.5\text{kHz}, \quad (2)$$

$$\tau_K(\phi) = \frac{2r\sin\phi}{c}, \quad < 2\text{kHz}. \quad (3)$$

Benichoux et al. [6] found the Woodworth formulation to work well for high frequencies, but at low frequencies, the Kuhn formulation [Eq. (2)] is more accurate.

*Correspondence should be addressed to: Aki Mäkivirta, Genelec Oy, e-mail: aki.makivirta@genelec.com

The ITD models of Woodworth and Kuhn describe only the horizontal plane. They are also of significantly different shape and differ for the same head radius particularly close to the 90° and 270° azimuths.

Larcher and Jot [7] and Savioja et al. [8] introduced the elevation into simplified spherical geometrical ITD models. The Larcher model for azimuth ϕ and elevation θ is

$$\tau_L(\phi, \theta) = \frac{r}{c}\,(\arcsin(\cos\theta \sin\phi) + \cos\theta \sin\phi), \tag{4}$$

and the Savioja model is

$$\tau_S(\phi, \theta) = \frac{r(\sin\phi + \phi)}{2c}\cos\theta. \tag{5}$$

These models do not include torso or shoulders, and the simplified spherical geometrical representation of the head is completely described with only the head radius. The ear channel entries are on the diametrically opposite sides of the sphere (antipodal).

These SH ITD models using only the head radius r are not linked to any single measurable dimension of the human head and assume non-realistic antipodal ear channel entry locations. Nevertheless, the head radius remains an important concept in the literature because it represents the general size of the head, linked to ITD magnitude.

Several methods have been suggested for obtaining relevant values for the head radius r. It is suggested that the head radius is taken to be the radius of a sphere having the same circumference as the head [9].

Algazi et al. [10] used the cardinal head dimensions (head height, width, and depth) in 25 subjects measured in photographs. They found linear combination coefficients of the head dimensions to produce the Woodworth head radius [see Eq. (1)], giving the best match to measured ITD. They called this the *optimal head radius*, and it is given as

$$r_o = 0.51w_{1/2} + 0.019h_{1/2} + 0.18d_{1/2} + 32 \text{ mm}. \tag{6}$$

This is defined using the half-width $w_{1/2}$, half-height $h_{1/2}$, and half-depth $d_{1/2}$ of the head.

Bomhardt et al. [9] used the Woodworth and Kuhn models with the optimal head radius from Algazi et al. [10] to find the head radius error that will cause audible differences in the ITD. This head radius error is shown to be more than 5–7 mm, depending on the direction of the sound source.

Sridhar and Choueiri [11] expanded the Algazi optimal head radius to elevation for the Woodworth and Kuhn models. They made the head radius a function of the elevation, achieving a slightly improved fit to the measured ITD without giving a calculation model. A similar effect was obtained by Romblom and Bahu[12] extending the SH by adding a block in the middle of the SH and changing dimensions of this block with elevation to fit the measured ITD better. Both [11] and [12] methods require knowledge of the full ITD before the model can be known, and they do not present a generally applicable model that could predict ITD for an individual prior to measuring the ITD.

Head width, depth, and height are the main factors determining the magnitude of ITD [5], and several authors [13–16] improved the SH model by considering these cardinal head dimensions. In fact, one aim of Romblom and Bahu was to use a head shape that better resembled the human head, and a more human-like head shape could be to model the head as an ellipsoid.

Studying the importance of the main features of the head, neck, and upper torso in contributing to the ITD and ILD, Cai et al. [17] found the SH model to be a fair simplification in describing the main features of the two, whereas an ellipsoid head shape had the highest improving effect. Duda et al. [13] described an ellipsoid head shape defined with the head width, depth, and height and also considered the ear positions, solving the sound propagation path geometrically. This more complex ITD model, in terms of the number of parameters, is reported to outperform the SH models.

Bomhardt and Fels [14] and Bomhardt et al. [15] described an ellipsoid having the width w and depth d of a head and then found a radius r_e as a function of the azimuth angle ϕ,

$$r_e(\phi) = \frac{w}{\sqrt{1 - \left(\frac{d^2 - w^2}{d^2}\cos^2\phi\right)^2}}. \tag{7}$$

Then they continued to find an analytic solution of the incident pressure and scattered pressure around a rigid sphere defined by this radius, finding HRIRs that then yield the times of arrival (TOAs) of sound for the left and right ears as the mean group delay in the frequency range 500 Hz–2 kHz. The difference of the TOAs is the ITD. Although the cardinal head dimensions spanning an ellipsoid form are used in the outset, this model finally describes the ITD using the SH shape.

Algazi et al. [10] and Bomhardt and Fels [14] disagreed about the relative importance of cardinal head dimensions. Whereas Algazi et al. found that the least important head dimension is the height, Bomhardt and Fels found it to be the depth. This demonstrates that there is no agreed mapping between the cardinal head dimensions and head radius used in spherical ITD models. However, both sets of authors considered head width, linked to the anthropometric interaural distance (AID), to be important.

Although the use of the cardinal head dimensions is an important step in bringing the SH model closer to reality, head dimensions are challenging to unambiguously measure and can therefore contain measurement inaccuracy, reducing their value in increasing the predictive accuracy of a model employing them. Models of the ITD assuming antipodal ear channel entry locations are front-back symmetric. Several SH models fall in this category. Measured ITD is not front-back symmetric, particularly for elevations above the horizontal plane, and this indicates the necessity to include non-symmetricity in a model predicting the ITD.

Aaronson and Hartmann [18] used an analytical solution of the scattered wave on an SH and demonstrated the Woodworth formula, which is defined on the horizontal plane, to predict the ITD well in high frequencies (>1.5 kHz). They presented extensions to Woodworth's model to include more realistic non-antipodal ear positions on the horizontal plane. Although they only described a model on

the horizontal plane, the ITD model proposed by Zhong and Xie [19] can also describe asymmetricity of ears.

Duda et al. [13] and Ziegelwanger and Majdak [16] described models that include representation of ear positions in the elevation and allowed using the SH model to predict ITD also in elevation. Ziegelwanger and Majdak [16] continued to use an SH model while allowing the ear positions on the head to be adjusted. They proposed two models for estimating the ITD using the TOA of the sound. The first model allows specific ear positions on the sphere relative to the center of the sphere. The second model also allows displacement of the head center away from the center of the coordinate system, typically defined by the measurement system layout. Using the first model, the direction-dependent path length on the ipsilateral side is

$$s_i = r_z\,(1 - \sin\theta_e \sin\theta - \cos\theta_e \cos\theta \cos(\phi_e - \phi)), \tag{8}$$

and on the contralateral side it is

$$s_c = r_z\,(1 + \arccos(\sin\theta_e \sin\theta + \cos\theta_e \cos\theta \cos(\phi_e - \phi)) - \frac{\pi}{2}\Big), \tag{9}$$

where the ear positions on the head are obtained by fitting the model to known ITD and given as azimuth ϕ_e and elevation θ_e of the ear for a head radius r_z. When the ipsilateral side is on the left hemisphere, the ITD becomes

$$\tau_z = \frac{s_i - s_c}{c}. \tag{10}$$

In addition to explaining the ITD using a simple geometry model, several alternative approaches to explain ITD are proposed. Brown and Duda [20] presented a direction-related decomposition of HRIRs recorded with real head measurements to explain the direction-dependent structure of HRIRs including azimuth and elevation of the sound source location. Although they present a model to explain the pressure events at the pinna, this approach does not offer a method of estimating values for the parameters needed to describe the HRIR model without first knowing a person's HRTF, making it difficult to use this model to predict the ITD for a person.

Another direct approach to describe the HRIR is given by Gamper et al. [21–23]. They used 3D head scans to simulate ITD based on the geometrically shortest acoustic wave propagation on the surface mesh describing a person's head. This method has the potential to produce an accurate ITD model because the estimate is derived using a precise 3D head data. Because detailed head geometry is required for the prediction process, this approach is not simple to use.

Zhong and Xie [19] fit the Fourier series to ITD data to predict ITD in azimuth at zero elevation once three anthropometric dimensions (tragial distance, pinna protruding height, and tragial backward circumference) are known. This model describes the individual characteristics as weights to the fundamental and second to fifth harmonic frequencies. The model exhibits front-back asymmetry and is based on a statistical fit to measured ITD data from 52 subjects, although the performance of the model is demonstrated on only four individuals.

The principal component analysis by Aussal et al. [24] uses a spherical harmonics model of the ITD and presents ITD variance with three orthogonal axes contained in the matrix of spatial basis functions $\mathbf{A}$ and weights $\mathbf{w}$ for each person and each axis:

$$\mathbf{A}\,\mathbf{w} = \tau. \tag{11}$$

Three principal components describe most of the variance. Correlating these components to physically meaningful anthropometric dimensions can establish the most relevant anthropometric features. Although the authors presented the linkage of the most significant principal components to anthropometric features, they did not present a prediction model on the level of an individual.

Approaches that study the HRIR in the time domain (Brown and Duda [20] and Gamper et al. [21–23]) or the frequency domain (Zhong and Xie [19]) or approaches that analyze the 3D shape of the ITD (Aussal et al. [24]) do not currently offer prediction of the ITD from easy-to-obtain anthropometrics. The SH models show potential for predicting the ITD from anthropometrics and have been improved to representing personal details, such as the head shape and ear location on the head.

The goal of the present paper is to study the feasibility of achieving personal ITD prediction with minimal input data about personal anthropometrics. Instead of using the abstract concept of the head radius, the input data used in this paper is the AID, a measurement of the direct geometric distance between the entries of the two ear channels. This measurement is related to the head width, observed to have the highest correlation with the maximum ITD value over other head dimensions [25]. Mapping between the maximum ITD and AID is presented, and then the SH ITD model is used to map the AID and ITD across the full range of azimuths and elevations. A local correction term is provided to improve the model fit. To do this work, four HRTF data sets comprising personal HRTF data of a total of 289 persons are used. The studied directions cover all the azimuth directions and –15° to 60° angles in elevation. Azimuth and elevation angles, coordinate system, and axis definitions follow *AES Standard 69:2015* [26], and angles are given in degrees.

HRTF results from anatomic features and shapes of the upper torso, head, and external ears. Wearing headphones therefore eliminates the HRTF. Processing audio signals with HRTF filters can recreate binaural 3D sound experience when headphones are used. ITD is a strong cue for localization and a significant feature to personalize in binaural processing [27].

Obtaining complete personal HRTFs can be difficult and time-consuming, and methods to produce individualized or personalized versions of general HRTFs or find the best likelihood fit in an existing HRTF in a database are popular [28, 29]. Personalization can involve first creating a minimum-phase presentation of the generic HRTF and then

adding the individualized personal ITD as a part of signal processing during binauralization [30, 27].

1 HRTF DATA SETS

The data used in the analysis in this paper consists of 119 measured and 170 computer-modeled HRTFs. The computer-modeled HRTF data set [31] includes 170 randomly selected persons, who gave their consent for research use of anonymized data. The data is produced using a photogrammetric method [32]. The method uses a video scan of the person's head and torso region. A 3D triangular mesh model of the person's external ears, head, neck, shoulders, and upper torso region is constructed and scaled to have the correct dimensions. Acoustic pressure fields are computed for a 2-m source distance with BEM software by solving the associated Helmholtz equation in the frequency-domain, similar to [33–36], for 836 directions in azimuth and elevation. The HRIR length is 1,024 samples and duration 21.3 ms. The sampling rate is 48 kHz. This data set will be referred to as DS1.

The measured personal HRTFs are adopted from the public-domain CIPIC, ARI, and ITA HRIR databases. The CIPIC HRIR database [37] contains 43 personal acoustically measured HRTFs taken with the closed meatus method. Of these, 37 subjects could be used as the data set in the analysis because they also had anthropometric data. A paraboloid surface is fitted close to the maximum ITD to find the maximum ITD values and their locations in the azimuth-elevation space. The HRIR data in CIPIC has a large gap close to the maximum ITD locations, and the paraboloid interpolation to find the maximum ITD location did not converge for four persons, so they were excluded. Thirty-three persons could be used in the data set. The HRTF is measured in 1,250 directions at 1-m distance. The measurements are taken with 44.1-kHz sampling rate, and the HRIRs are windowed with short Hanning window tapers at the start and end of the captured HRIR to remove room reflections in the measurements. The resulting HRIRs are 200 samples in length and have a 4.5-ms duration. This data set is referred to as DS2.

The ARI HRIR database [38] contains 44 subjects with full anthropometric details, and all of these could be used. The HRTFs are measured in a semianechoic room with the closed meatus method in 1,550 directions at 1.2 m. The horizontal and vertical range is 360 and –30 to 80, respectively. The sampling frequency is 48 kHz. The head orientation is monitored to ensure accuracy before acquiring each measurement. The impulse responses are windowed with asymmetric Tukey window with 0.25-ms onset and 1-ms fade out at the end to a 5.33-ms duration. This data set is referred to as DS3.

The ITA HRIR database [39] contains 48 subjects, with a set of 42 subjects who could be used in this work. Three subjects did not have the full set of data and could not be included in the analysis. Three subjects were excluded because of problems in the continuity of the ITD data and issues in the sampled impulse responses. The measurement setup consists of 64 loudspeakers at the radius of 1.5 m with a 2.5° spacing from 1.55°–160° zenith in a semi-anechoic chamber, resulting in 2,304 directions. Closed meatus impulse responses are cropped to 256 samples. The sampling frequency is 44.1 kHz. This data set is referred to as DS4.

These data sets come from varied types of adult populations, contain both male and female listeners, include both acoustically measured and numerically modeled HRIRs, and therefore offer a good basis for understanding how generally the ITD prediction model can cover adult population with typical anthropometric variation.

2 AID

In this work, AID is used as the descriptor of the person's ITD. Traditionally the concept that is used for describing the effect of the head size to ITD is the head radius; however, head radius cannot be directly measured on a person. Mostly the head radius is calculated (backward) as the *effective head radius*, the value that yields the measured ITD. Methods to predict the value that should be used as the head radius for humans have been suggested.

The AID is estimated in DS1–DS4 in the following ways. For the DS1, the AID is computed as a the direct distance between the ear channel entries, measured in the 3D geometry data of each person. The performance of the 3D video scanning method applied in DS1 is studied by comparing the 3D model generated using the video scan method of the Kemar mannequin to the mannequin manufacturer's 3D data [32]. The distance between the ear channel entries in the 3D video scan (130.1 mm) agrees well with the manufacturer 3D data (129.3 mm) in the Kemar head.

For the DS2 data, the AID is not directly available, and the published information [10] on the DS2 data does not give full disclosure of how the anthropometric measurements were taken. The exact method of obtaining the "head width" (x_1) is not explained and appears to measure the maximum width of the head. This work shows that this is not the same dimension as the AID because the ear channels open into cavum conchas, and direct measurements of the AID address the distance at the cavum concha bottom where the ear channel opens to the skull. In [10], "cavum concha depth" for both ears d_8 and d_{16} has been given. Only a part of this value is the radial depth of the cavum concha. As the cavum concha is typically slightly inside the skull surface, to obtain the reasonable estimate of the AID, a combination of the "head width" (x_1) and "cavum concha depth" d_8, d_{16} measurements are used:

$$2r_a = x_1 - \left(\frac{d_8}{2} + \frac{d_{16}}{2}\right). \tag{12}$$

For DS3, the anthropometric data reported contain measurements of the head width and concha depths for each person, defined in the same way as for DS2. For DS4, the head width value comes from the MRI image data for each person, and the concha depths are not available, so the mean of the pooled concha depth data from data sets DS2 and DS3 is applied. The mean concha depths for the data sets DS2 (9.8 mm; N = 33) and DS3 (13.6 mm; N = 44) are slightly different. The pooled mean becomes 12.2 mm.

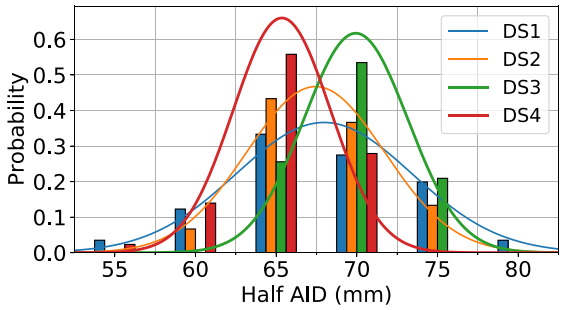

Fig. 1. Anthropometric interaural distance (AID) distributions in data sets (DS1–DS4). Note that data in each 5-mm AID span is represented side-by-side for the data sets for sake of clarity. The histograms are calculated for 5-mm bin widths, located at the same center values for all data sets. The bins at the same center value are grouped together, with the center value indicated on the horizontal axis.

Table 1. Anthropometric interaural distance (AID).

Data set	N	Mean (mm)	Median (mm)	SD (mm)
DS1	170	136	136	10
DS2	33	136	136	8
DS3	44	140	140	6
DS4	42	130	132	6
All	289	136	135	10

The distributions of the AID for the data sets shows differences that are related to the methods of measuring head width and concha depth and variation of the population in the samples (Fig. 1; Table 1).

3 MAXIMUM ITD MAGNITUDE AND DIRECTION

Several methods are used for estimating the ITD from a set of HRTFs. Katz and Noisternig [40] and Andreopoulou and Katz [27] presented reviews of the most common ITD estimation methods. ITD estimators that operate on low-pass–filtered HRIR are reported to have good match to subjectively estimated ITD. In preparation for this study, several well-performing methods reported in [27] were evaluated for accuracy and robustness across all the data sets.

When the HRIR low-pass–filter corner frequency reduces, the ITD estimated by a threshold detector changes systematically (Fig. 2), making the correct ITD value ambiguous. Additionally, the measured HRIR data contains inherent measurement noise preventing the use of low-threshold values.

Because low-pass filtering prior to estimating ITD can contribute a systematic change (Fig. 2), low-pass filtering is not considered suitable for this analysis of the value and location of the maximum ITD. Uncertainty regarding the ITD value could be avoided by using a cross-correlation method and no low-pass filtering. Cross-correlation shows less systematic change of the ITD estimate with changing low-pass corner frequency, and the estimated value remains

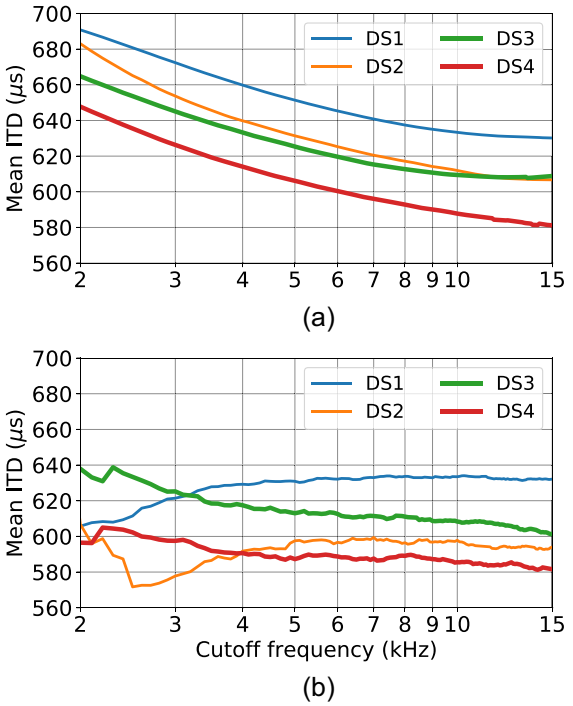

Fig. 2. The effect of low-pass filtering to the interaural time difference (ITD) value. The ITD value shown is the mean across all persons. The HRIR closest to the direction azimuth 80°, zero elevation, is low-pass–filtered, and the ITD is detected using the threshold method at –15-dB threshold level (a) and the cross-correlation method (b). The low-pass filter is a fourth-order maximally flat magnitude filter with linear phase. DS = data set.

relatively stable when the HRIR bandwidth is not severely limited.

The cross-correlation method originally described by Nam et al. [41] finds the maximum cross-correlation between an unfiltered HRIR and its minimum phase version, individually for both ears, and takes the difference in the lags for the correlation maxima in the left and right ears as the ITD value. This method is called "PhminXcorr" in [27], and an implementation of this is available as a MATLAB code in [42].

When HRIR low-pass filtering is not used, the Nam et al. ITD detector is considered the best after the interaural cross-correlation (IACC) method for the normalized sum of success rates at one just-noticeable difference (JND) level [27]. However, it was observed that the IACC method does not produce reliable ITD estimation when the audio source is located laterally. The Nam et al. method is less susceptible to correlation lag determination errors when correlating between left and right ear HRIRs, unlike the IACC method [27].

The maximum ITD values and their locations in the azimuth-elevation space are found by fitting a paraboloid surface to the ITD values close to the maximum. The parabolic surface is fitted in the range 75°–105° or 255°–285° in azimuth and –10° to 20° in elevation. A weighting is applied such that the highest weight is assigned closest to the maximum of the ITD values.

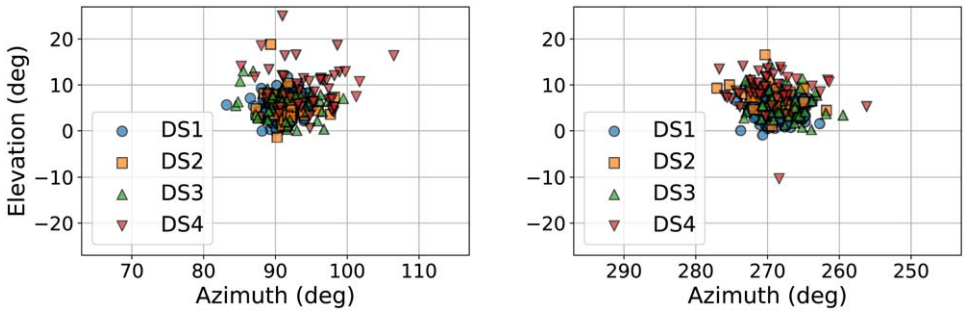

Fig. 3. The maximum interaural time difference (ITD) magnitude locations for each subject (DS1, N = 170; DS2, N = 33; DS3, N = 44; and DS4, N = 42). DS = data set.

The locations of the pooled mean maximum ITD across all the data sets is 91.4° azimuth and 4.7° elevation for the left ear and 268.6° azimuth and 5.2° elevation for the right ear. The four data sets agree well on the directions of the maximum ITD (Fig. 3).

The angular sampling intervals in the azimuth and elevation directions differ between the data sets. The data sets generally do not contain a measurement in the direction of the maximum ITD. For example, DS2 has a directional sampling plan that has a 20° quantization of the angle close to 90° and 270° directions. A 2D paraboloid fit is used in the vicinity of the anticipated location of the ITD maximum, and this also helps to handle outliers and noise in the ITD value data. This approach enables comparison of the maximum ITD data across data sets. In a few cases, it was not possible to find an estimate of the maximum ITD, and these persons were excluded, as explained in SEC. 1. The four data sets show similar distributions of the ITD, with DS2 showing smaller maximum ITD values than other data sets (Fig. 4; Table 2).

Table 2. Maximum interaural time difference (ITD) value and its location (mean ± SD).

Data set	ITD (μs)	Azimuth (degree)	Elevation (degree)
Left ear			
DS1	669 ± 42	90.9 ± 1.8	5.3 ± 2.1
DS2	637 ± 33	91.8 ± 2.5	5.0 ± 3.4
DS3	659 ± 36	91.4 ± 3.5	6.0 ± 3.0
DS4	657 ± 29	94.6 ± 4.3	9.6 ± 4.2
Right ear			
DS1	666 ± 42	269.0 ± 2.1	5.1 ± 2.1
DS2	638 ± 27	269.2 ± 3.2	7.4 ± 3.0
DS3	651 ± 35	267.9 ± 3.2	5.8 ± 3.3
DS4	665 ± 34	269.6 ± 4.0	8.7 ± 4.0

4 MAPPING FROM AID TO ITD

Data sets in this work enable comparison of the ITD predicted using the SH model and half of AID in place of the head radius to the ITD estimated from HRIR data. A multiplier can now be determined to match the ITD in a Ziegelwanger and Majdak SH model [16] on the horizontal plane for each individual to the actual ITD of the individual using least-squares fitting. The distribution of this multiplier is presented for all the persons in all the data sets (Fig. 5). The grand mean ± standard deviation value of the multiplier across all data sets is 1.30 ± 0.06.

Next, the relationship between the AID and maximum ITD value and direction is studied. A model is created to

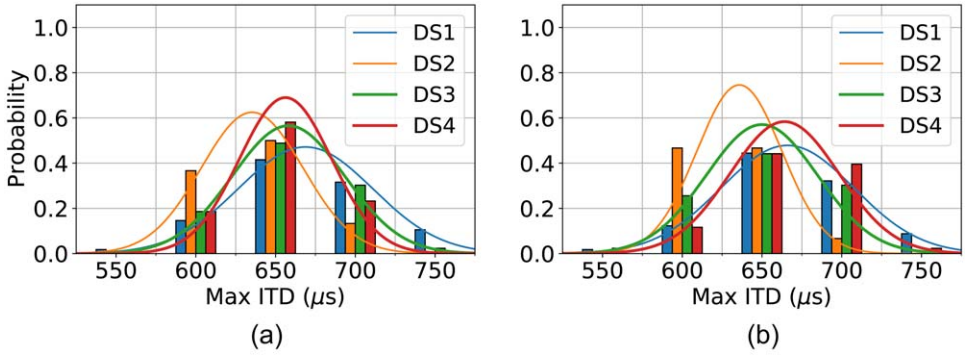

Fig. 4 Distributions of the maximum interaural time difference (ITD) magnitude for the positive maximum (a) and for the negative maximum (b) on DS1 data (N = 170), DS2 (N = 33), DS3 data (N = 44), and DS4 data (N = 42). The normal distributions with the same mean and SD as in the data set are also depicted. The histograms are calculated for 50-μs bin widths. See Fig. 1 for explanation of the grouping of bins.

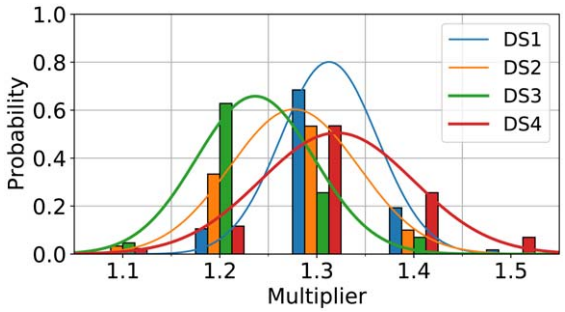

Fig. 5. The least-squares fit at zero elevation to find the multiplier for each person's anthropometric interaural distance (AID) to match the spherical head model interaural time difference (ITD) prediction to the actual ITD data when half of the AID is used as the head radius. The bins at the same center value are grouped together, with the center value indicated on the horizontal axis. The mean and median multipliers are 1.31 for data set DS1 (N = 170), 1.28 for DS2 (N = 33), 1.24 for DS3 (N = 44), and 1.32 for DS4 (N = 42).

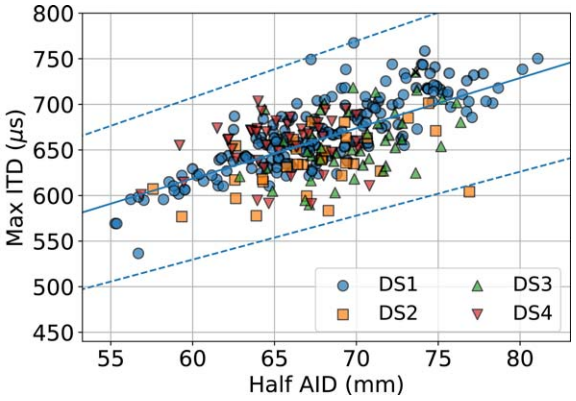

Fig. 6. Maximum interaural time difference (ITD) correlation to the anthropometric interaural distance (AID) for pooled DS1 (N = 170), DS2 (N = 33), DS3 (N = 44), and DS4 (N = 42) data. The least-squares fit regression line [Eq. (13)] is shown with 95% confidence limits. DS = data set.

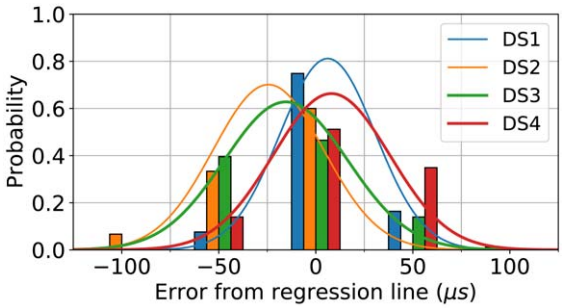

Fig. 7. The error of predicting the personal maximum interaural time difference (ITD) using the regression model and the personal anthropometric interaural distance (AID). The deviation (error) of the individual maximum ITD from the regression line is shown. The mean ± SD values are 6.2 ± 24.5, −23.8 ± 28.0, −14.0 ± 32.5, and 8.2 ± 30.4 for DS1, DS2, DS3, and DS4, respectively. The normal distributions are fitted for reference. The histograms are shown for 50-μs bins. See Fig. 1 for explanation of the grouping of bins. DS = data set.

predict the maximum ITD value when the AID ($2r_a$) is known. The coefficients $\{k, m\}$ are found for a regression model such that this regression model has the best least-squares fit to the estimated maximum ITD max(τ) observed for all the persons in the pooled data sets DS1–DS4:

$$\max(\tau) = k \cdot r_a + m. \tag{13}$$

The least squares estimates can be used to obtain a prediction interval to assess the credibility of each $[r_a, \max(\tau)]$ observation. The prediction interval can be computed as the product of the predicted standard deviations of each estimate and Student's t value of the sample data. The standard deviations are obtained from the covariance matrix of the estimates, which is given as

$$\Sigma = \sigma_e^2 (H^T H)^{-1}, \tag{14}$$

where H is the observation matrix for the linear regression model and σ_e^2 is the variance of the residuals of the fitted model. The diagonal of Σ yields the variances $\{\sigma_{\hat{k}}, \sigma_{\hat{m}}\}$ of the estimated coefficients $\{\hat{k}, \hat{m}\}$, respectively. The credibility limits for the estimates are

$$k = \hat{k} \pm t_N \sigma_k, \tag{15}$$

$$m = \hat{m} \pm t_N \sigma_m. \tag{16}$$

t_N is the Student's t value for sample size N.

This results in coefficients $k = (5.53 \pm 0.69) \cdot 10^{-3}$ and $m = (286.63 \pm 47.29) \cdot 10^{-3}$ at 95% confidence level for pooled data. The resulting fit on pooled data is shown in Fig. 6. The resulting residual after applying the regression (Fig. 7) demonstrates the power of the regression model to detect the maximum ITD. The actual data deviates from the regression under 100 μs in all data sets.

5 EXPLANATORY POWER OF THE SH MODEL

This work aims to create an ITD prediction model with a single anthropometric variable. The SH model by Ziegelwanger and Majdak [Eq. (10)] covers the full sphere of sound arrival directions and allows the ear positions to be presented, representing well the capability of SH models to describe the ITD. Next, an analysis of the explanatory power of this SH model is presented.

A scaling factor adjusts the AID dimension. Half of the AID is used in place of the head radius r_z for the SH model in Eq. (10). The average scaling factor is used for each data set DS1–DS4. The scaling factors are the mean values seen in Fig. 5 for DS1–DS4. Additionally, the ear locations are set using the average maximum ITD directions for each ear and data set, which are given in Table 2. Several SH models place the ears antipodally, exactly at 90° relative to the forward direction. The Ziegelwanger and Majdak model used in this work assumes that the maximum ITD occurs when the source of audio is at the opposite direction to the contralateral ear direction, so there is a relationship between the direction of the ear and of the maximum ITD.

When $\tau_{e,i}$ is the ITD estimated from the HRIR data and $\tau_{m,i}$ is the ITD model prediction of the ITD for a subject i,

the prediction error for the individual is $\Delta\tau_i$, and the mean prediction error $\bar{\varepsilon}$ across all persons becomes

$$\bar{\varepsilon}(\phi, \theta) = \frac{1}{N}\sum_{i=1}^{N} \Delta\tau_i = \frac{1}{N}\sum_{i=1}^{N} \tau_{e,i} - \tau_{m,i}, \tag{17}$$

where the total number of subjects in a data set is N.

Large local prediction errors are observed in the SH ITD model. On average, the SH ITD model has a challenge predicting ITD correctly particularly in two regions, located symmetrically in azimuth and at similar elevation, left and right at $(\phi, \theta) = (80°–120°, 10°–60°)$ and $(\phi, \theta) = (240°–280°, 10°–60°)$ (see Fig. 8). A 50-μs limit is also depicted to help with understanding where the large deviations between the SH prediction and actual ITD value are located. All data sets DS1–DS4 show similar characteristics. The main finding here is that the systematic deviation is in two main directions. To account for the local errors, a correction can be introduced for the poorly modeled directions of Eq. (10).

In Fig. 8(c), the ITD prediction error data shows the variation along the elevation occurring in synchrony with the vertical angular sampling. The data is sampled at 5° elevation intervals. The variation effect appears to be caused by the left-right impulse response pairs containing a timing difference in the HRIR data of one sampling period (about 20.8 μs) back and forth in time with changing elevation, and then the consecutive elevations show the variation seen in the ITD data.

Also, in Fig. 8, the prediction error data of DS4 shows an anomaly close to the +15° elevation. This is produced by the structure of the HRIR data in this elevation, where the impulse responses show an unexpected early feature in the contralateral ear HRIR at 15° elevation, which is not seen in higher or lower elevations.

6 ENHANCED SH MODEL

To create an enhanced model, the authors look at DS1. The non-linear least squares fit is performed across all persons and in all selected directions to the SH ITD model, with a factor added to introduce local improvement of the prediction in specific directions where the SH ITD model has poor performance.

The enhanced ITD model uses a correction term defined as a bivariate density function. This is optimized to minimize the maximum error within the two regions in which the 50-μs deviation limit is exceeded. The enhanced model to estimate the ITD $\tau_m(a, \phi, \theta)$ across the azimuth ϕ, elevation θ, and AID a, is given as

$$\tau_m(a, \phi, \theta) = b_1\tau_Z(a/2, \phi, \theta) - b_2[P_1(\phi, \theta) + P_2(\phi, \theta)], \tag{18}$$

where $b_1 = 1.31$ is the multiplier for the Ziegelwanger and MajdakSH model (Eq. 10) and $b_2 = 0.16$ is the multiplier for the correction term. The coefficient b_1 is obtained as the average least-squares fit at zero elevation to the AID DS1. This coefficient can also be seen in Fig. 5. DS1 is used in determining the model because DS1 contains the largest number of individuals. The resulting model is then

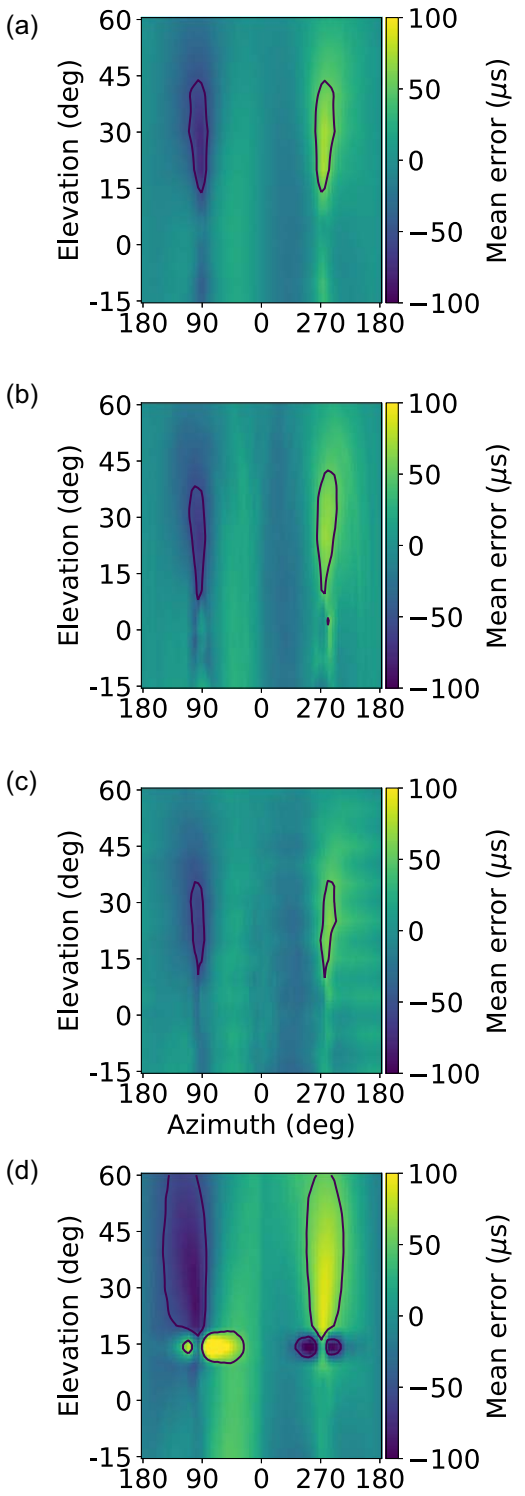

Fig. 8. The mean prediction error $\bar{\varepsilon}$ of the match of the observed interaural time difference (ITD) to the spherical head (SH) ITD model prediction τ_z. The SH model is scaled using the data set specific mean multipliers given in Fig. 5, and the data set specific mean ear positions on the head are given in Table 2. Black lines indicate the 50-μs level. Data sets from top to bottom are DS1 (N = 170) (a), DS2 (N = 33) (b), DS3 (N = 44) (c), and DS4 (N = 42) (d).

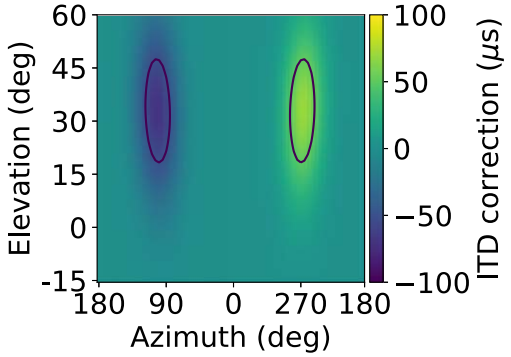

Fig. 9. Correction term for the spherical head (SH) model. Black lines indicate the 50-μs level.

applied to the other data sets to study how well this model can predict ITD for other HRIR data.

The bivariate density functions P_i are defined as

$$P_i(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^2 \det \Sigma_i}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1} (\mathbf{x}-\mu_i)\right], \tag{19}$$

where $i = \{1, 2\}$ and $\mathbf{x}$ is the vector expressing the location in the azimuth and elevation:

$$\mathbf{x} = [\phi\theta]. \tag{20}$$

The mean of the value range to be corrected (correction center location) is set by μ_i and the covariance matrix Σ_i describes the correction shape, both appearing in the azimuth ϕ and elevation θ. The coefficients for these local corrections obtained with the least-squares fit (angles expressed in degrees) are shown below, and the correction term $b_2(P_1(\phi, \theta) + P_2(\phi, \theta))$ is shown in Fig. 9.

$$\Sigma_1 = \begin{bmatrix} 425.6 & 35.9 \\ 35.9 & 328.6 \end{bmatrix}, \tag{21}$$

$$\Sigma_2 = \begin{bmatrix} 425.6 & -35.9 \\ -35.9 & 328.6 \end{bmatrix}, \tag{22}$$

$$\mu_1 = \begin{bmatrix} 99.3 \\ 33.6 \end{bmatrix}, \tag{23}$$

$$\mu_2 = \begin{bmatrix} 260.7 \\ 33.6 \end{bmatrix}. \tag{24}$$

In the following, these model parameters are applied to DS1–DS4. The mean prediction error (Fig. 10) indicates that the systematic error reduces significantly for all data sets when the correction term is applied. On average, the prediction of the enhanced ITD model matches the observed ITD both in the computer-modeled (DS1) and measured HRIR data (DS2 and DS3) better than ±70 μs in all azimuths for positive elevations up to 60°, covering the essential application range well. This is also true for DS4 except at the 15° elevation region. Note that the model is derived using the DS1, so it should be expected that DS1 shows the best fit.

The maximum error in DS1 across all persons is less than 189 μs for each direction. The maximum prediction error

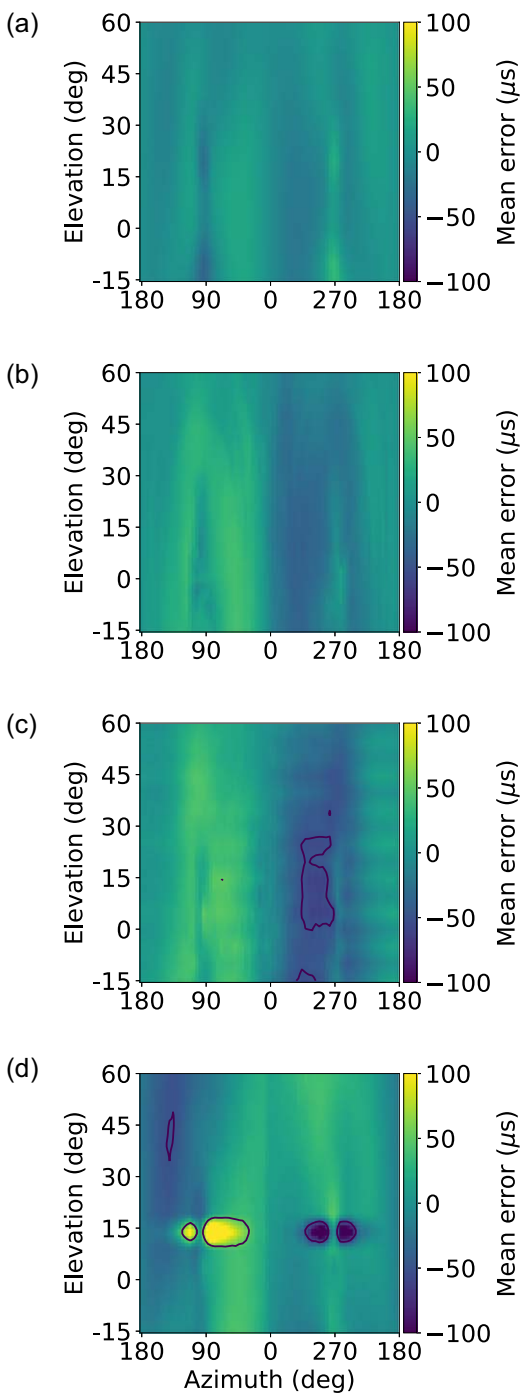

Fig. 10. The mean prediction error $\bar{\varepsilon}$ between the interaural time difference (ITD) data and the enhanced spherical head ITD model given in Eq. (18). DS1 (a), DS2 (b), DS3 (c) and DS4 (d). Black lines indicate the 50-μs level. DS = data set.

occurs at 10° elevation and 100° azimuth for left hemisphere and 0° elevation and 105° azimuth for right hemisphere. Similarly for DS2, the maximum error is less than 271 μs, and the maximum error directions are (100, –3) and (260, 2). The maximum error occurs around the focal point of the measurement coordinate system having the least density of HRIR measurements. For DS3, the maximum error is less than 242 μs, and the maximum error directions are (115,

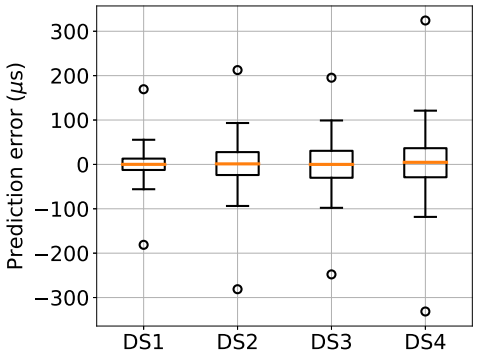

Fig. 11. The distribution of the enhanced interaural time difference (ITD) model prediction error in each data set across all persons, all azimuth angles, and elevation angles in the range of –15° to 60°. The median and mean of the prediction error coincide and are shown as the orange line inside the box. The SD of the prediction error is represented by the box edges. The prediction errors at the first and 99th percentiles in the cumulative distribution of the data set are indicated by whiskers, and the values for these are –55.9 and 55.5 μs for DS1, –91.1 and 92.1 μs for DS2, –97.5 and 98.8 μs for DS3, and –118.5 and 121.8 μs for DS4. The minimum and maximum of the prediction error are indicated by round outlier marks.

0) and (250, 0). The maximum deviation occurs extremely locally at the direction given. For DS4, the maximum error is less than 320 μs, and the maximum directions are (75, 14) and (280, 14). Because these maxima appear to be associated more with the local characteristics of the collected acoustical data, they may not be taken as indications of the worst prediction quality of the enhanced SH model.

To give an idea of the statistics of the prediction error for each data set, a box-and-whiskers presentation of the error is given in Fig. 11. Although the mean and median errors are close to zero and the SD of the error is small, there is a small fraction of larger errors. The outliers with the prediction error larger than the 99th or first percentile were studied separately.

The first reason for outliers is that for certain persons the AID was not sufficient to predict the ITD, and the prediction was systematically either too small or too large. These persons can be seen as outliers also in Fig. 6. The second main reason for outliers were persons for whom the direction of the maximum ITD was far from typical. These cases can also be seen as outliers in Fig. 3. For these persons, there is a systematic mismatch to the model as the maximum ITD direction of the model is fixed and is not personalized.

The spatial direction of these largest errors is studied in Fig. 12. The largest proportion of persons exceeding the first and 99th percentile limits are seen close to the 90° and 270° azimuth for which the JND of the ITD difference is also known to have large values.

Although the enhanced model correction term that has been determined using photogrammetrically obtained ITD data, the correction term also shows reasonable performance with acoustically measured data (Figs. 11, 12, and 13).

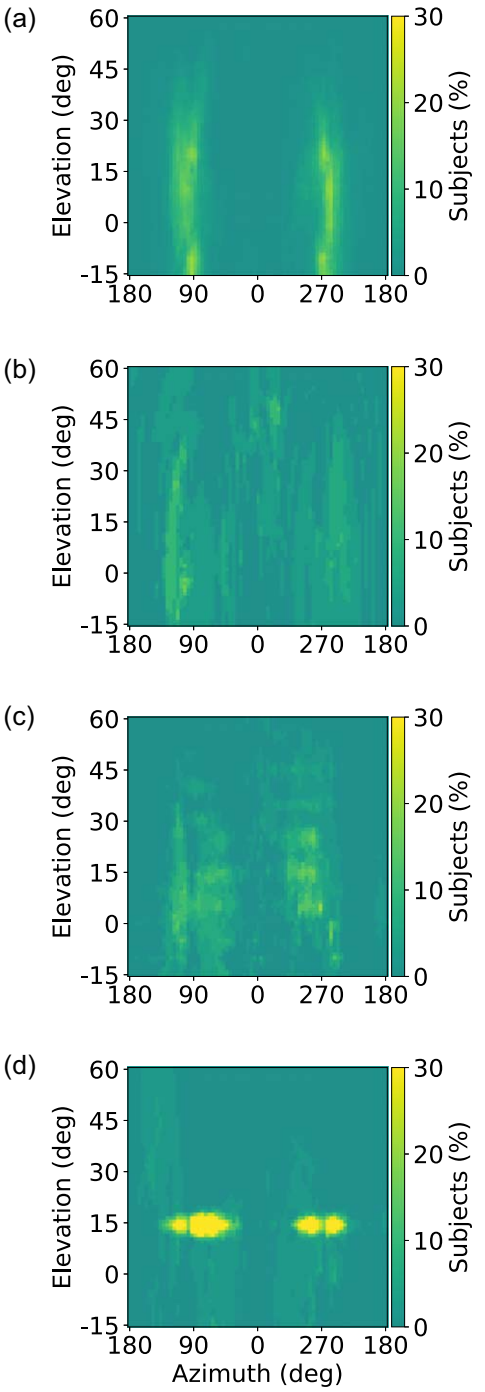

Fig. 12. The relative number of subjects in each azimuth-elevation direction below the first or above the 99th percentile level shown in Fig. 11. Data sets from the top down: DS1 (a), DS2 (b), DS3 (c), and DS4 (d).

The ITD prediction error can be positive when the actual ITD value is larger than the prediction or negative when the prediction is larger than the actual ITD value. The proportion of persons with the maximal prediction error across all studied directions under a certain limit is given in Fig. 13. For DS4, the data anomaly at elevations 10°–20° has been excluded here. This ITD error level can be compared to known JND of the ITD to understand the likelihood of

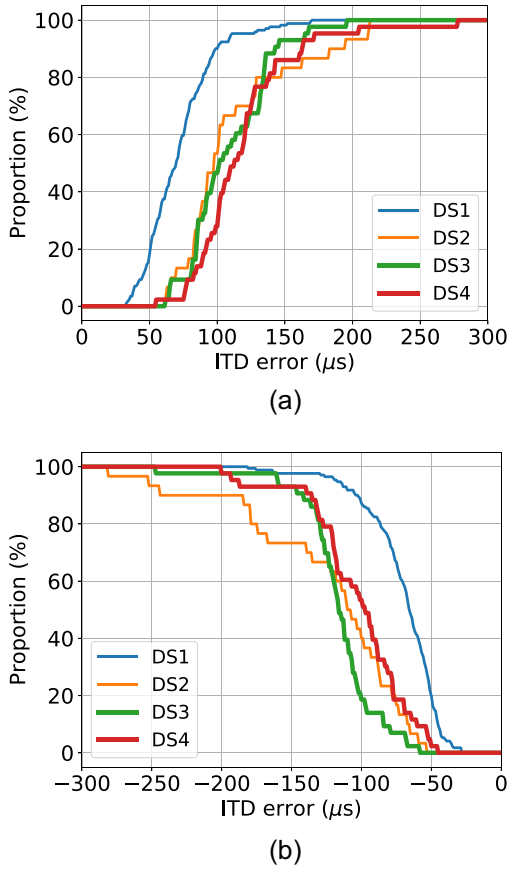

Fig. 13. The proportion of persons with the magnitude of the maximum interaural time difference (ITD) error in all directions under the ITD error given on the horizontal axis (a) for the positive ITD error and (b) for the negative ITD error. For DS4, the data anomaly at elevations 10°–20° has been excluded. DS = data set.

audibility, although the challenge is that the JND of ITD change is only known in very few directions and that only the maximal error across all studied directions is shown. Simon et al. [30] reported the median JND for ITD at zero elevation and 90° azimuth as 68–125 μs and at 30° azimuth as 24–44 μs using three different measurement methods. Bomhardt et al. [9] considered JND at zero elevation in the azimuth range 345°–270° and reported the median JND of ITD for 32 persons at 270° azimuth direction as 80 μs and at 330° as 30 μs. Andreopoulou and Katz [27] reported the median JND at 90° as 109.9 μs and 30° as 72.5 μs. The difference between this work and [9] is suggested to be because of differences in the methodology.

7 DIRECT MEASUREMENT OF AID

Using AID for predicting ITD would be particularly interesting if it was possible to directly measure AID on a person. To evaluate the feasibility of this, the AID is measured mechanically for nine persons in DS1. For this, a gauge with two 11.2 mm diameter straight-end dowels is used. The gauge is closed such that the dowels touch lightly at the bottoms of the left and right ear cava concha. The gauge position reading is noted, the gauge is removed, and then readjusted to the same reading. The distance between the dowels is recorded as the mechanical AID. This dowel diameter is suitable to reliably touch the cavum concha floor, a structure typically flush with the ear channel opening, and not too small such that the dowel would sink inside the entry of the ear channel. Each person to be measured moves the arms of the measurement jig. The measurement on each person is repeated two to three times.

The photogrammetric ear channel entry points are determined in the photogrammetric 3D models. The distance between these points is taken as the photogrammetrically determined AID. The mechanically measured AID is on average 1.0 mm larger than the photogrammetrically estimated AID. The SD across persons in this data set is quite large, 6.0 mm, whereas the typical SD of repeated measurements for an individual person is 0.8 mm. To put this in context, 1-mm difference in the AID translates to a about 4.6-μs time difference in the ITD at the maximum. The sample of persons in this test is small but serves to demonstrate direct measurement of the AID. The large variation across persons in this experiment appears to be related to the seating process with the gauge. The seating is determined by each person based on how the dowels were touching the ears. Discussing with the persons, there were variations in how people interpreted the instructions of "light touch." This method does not use an objective means to determine the contact, such as a pressure switch, and including an objective indicator may improve the measurement similarity across persons.

8 DISCUSSION

SH models traditionally describe the ITD using the concept of the head radius as the input, but the head radius cannot be directly measured on a person. Methods to predict the head radius using measurements of the head width, height, and depth have been proposed considering the ITD at zero elevation [10, 15]. These find the head width to be the most significant factor in determining the head radius value. This motivates this study using the AID for predicting the ITD.

The photogrammetrically obtained HRIR DS1 and acoustically measured HRIR DS2–DS4 display similar characteristics of the ITD. All data sets DS1–DS4 include some anthropometric information, but the exact methods of measurement and selection of anthropometric data differ slightly, making direct comparisons more difficult. The participants in the data sets are adults, with the AID ranging from 110 to 164 mm. The mean AID across all the data sets is 136 mm.

If there is an error in measuring the AID on a person, this results in an error in the ITD, which can become audible when the JND is exceeded. The maximum error occurs close to 90° and 270° azimuth, at zero elevation. Then, a 1-mm measurement error produces a 4.6-μs error in the ITD. When the elevation increases or decreases, this error sensitivity reduces.

The ear channel opens to the concha. The reasonable estimate for the mean concha depth is found to be 12.2 mm,

but there is significant individual variation. The correlation between the head width and concha depth in each person does not show correlation (Pearson correlation is 0.26 for DS2 and –0.03 for DS3). Difference in the left and right ear concha depths is not significant; the concha depth difference mean is 5.5% (DS2: SD 21%; N = 43) of the personal left and right concha depth mean.

This material does not allow analyzing the significance of the combination of the cardinal head dimensions instead of using only the AID as the input data. On the other hand, the AID may be easier to measure systematically in the same way as the point in which the ear channel opens to the concha, whereas methods for head width measurement are somewhat lacking a clear definition. For example, Bomhardt et al. [15] measured the head width with a caliper as the distance between the processus temporalis of the cheekbones.

All the data sets agree well in terms of the direction and magnitude of the maximum ITD, but comparing the data sets requires the use of interpolation because the spatial sampling close to maximum ITD differs significantly between the data sets, because the angular quantization close to the maximum ITD can be large (DS2), and because there can be significant variability in the measured ITD data particularly close to the maximum ITD. Directly picking maximum data values would lead to significant errors both in the maximum ITD values and their directions. The 2D interpolation using a paraboloid surface fitted in the ITD data in the vicinity of the maximum ITD offers reasonable estimation of the values and directions of the maximum ITD and allows comparisons between data sets. Some authors have used spherical harmonic model fits to HRTF data (see, for example, [11, 43]), and spherical harmonics could, in principle, fit to the complete 3D ITD profile, but these methods easily become computationally heavy without offering significant benefits over local fitting close to the maximum ITD value for the type of problem being worked with.

Algazi et al. [10] and Bomhardt et al. [15] studied calculation of the head radius. They treated the head radius as a constant value, tightly linked to the SH concept, and they mainly looked at the correct way to combine the various cardinal head dimensions to produce the constant radius value with the best fit. Shridhar et al. [11] considered the contributions of the head dimensions to the value of the head radius, allowing it to vary as a function of frequency, but did not make the head radius itself a function of direction.

Bomhardt et al. made a similar comparison as the current authors do between the ITD model and measured ITD data ([15]; Fig. 8; TOA method), although they did not report a feature similar to the current authors' local correction. They used a steep low-pass filter with the corner set at 2 kHz before detecting the ITD in the HRIR data. After a similar filter is applied to HRIR data, the local variation reported is no longer visible, demonstrating the effects of applying low-pass filtering prior to detecting the ITD.

The local enhancement of the accuracy of SH models in higher elevations for a certain range of directions is demonstrated to be needed for all the four independent data sets, which are created by two different methods (simulation of acoustics on photogrammetrically modeled geometry and measured acoustically) and contain a total of 289 individuals. All the data sets contain the local feature requiring the enhancement, although in DS4 the local feature is showing in a more pronounced way. It is also anticipated that similar features may exist at negative elevations, but these data do not allow further analysis of this.

The enhanced SH model is computationally light and suits personalizing ITD in real-time applications, such as for virtual reality or gaming. Binauralization tasks frequently use general HRIRs obtained from a head-and-torso simulator, or HRIRs are obtained with a best-fit selection process in a library of measurements. These HRIRs are then personalized for best acceptability. The proposed enhanced SH model can be applied to personal ITD predictions in such binauralization applications. Models describing the time-domain structure of the HRIR [44, 20] allow ITD personalization because these models represent the ITD effects explicitly.

9 CONCLUSION

This paper proposes an enhanced SH model describing personal ITD for a wide range of sound arrival directions. A photogrammetrically obtained personal HRIR data set is employed in generating the enhanced SH model. Then, three acoustically measured HRIR data sets are used in evaluating the performance of this model. The enhanced SH model uses the personal AID, the direct distance between the ear channel entries, instead of the traditional concept of the head radius. The proposed local enhancement term for the SH model improves the accuracy for elevation angles above the ear level. The first and 99th percentile levels of the ITD prediction error for all persons and in all directions remain below ± 122 μs.

10 ACKNOWLEDGMENT

The DS2, DS3, and DS4 HRIR data bases are publicly available. For DS1, although the users of the Genelec Aural ID service have agreed to scientific research use of pooled non-identifiable data, there is no consent from individuals to make the data publicly available. The authors express their gratitude to all persons appearing in the data sets and to the institutions allowing access to the public data sets.

11 REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).

[2] V. Pulkki and M. Karjalainen, *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics* (Wiley, West Sussex, UK, 2015).

[3] F. L. Wightman and D. J. Kistler, "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization," *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1648–1661 (1992 Mar.). https://doi.org/10.1121/1.402445.

[4] R. S. Woodworth and M. Schlosberg, *Experimental Psychology* (Holt, Rinehart and Winston, New York, NY, 1958).

[5] G. F. Kuhn, "Model for the Interaural Time Differences in the Azimuthal Plane," *J. Acoust. Soc. Am.*, vol. 62, no. 1, pp. 157–167 (1977 Jul.). https://doi.org/10.1121/1.381498.

[6] V. Benichoux, M. Rébillat, and R. Brette, "On the Variation of Interaural Time Differences With Frequency," *J. Acoust. Soc. Am.*, vol. 139, no. 4, pp. 1810–1821 (2016 Apr.). https://doi.org/10.1121/1.4944638.

[7] V. Larcher and J.-M. Jot, "Techniques d'Interpolation de Filtres Audionumériques: Application à la Reproduction Spatiale des sons sur Écouters," in *Proceedings of the Congrés Francais d'Acoustique (CFA)* (Marseille, France) (1997 Apr.). https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.1017.

[8] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating Interactive Virtual Acoustic Environments," *J. Audio. Eng. Soc.*, vol. 47, no. 9, pp. 675–705 (1999 Sep.).

[9] R. Bomhardt, I. Patino Mejía, A. Zell, and J. Fels, "Required Measurement Accuracy of Head Dimensions for Modeling the Interaural Time Difference," *J. Audio Eng. Soc.*, vol. 66, no. 3, pp. 114–126 (2018 Mar.). https://doi.org/10.17743/jaes.2018.0005.

[10] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a Spherical-Head Model From Anthropometry," *J. Audio. Eng. Soc.*, vol. 49, no. 5, pp. 472–479 (2001 Jun.).

[11] R. Sridhar and E. Y. Choueiri, "Capturing the Elevation Dependence of Interaural Time Difference With an Extension of the Spherical-Head Model," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), paper 9447.

[12] D. Romblom and H. Bahu, "Blockhead: A Simple Geometric Head Model," in *Proceedings of the AES International Conference on Headphone Technology* (2019 Aug.), paper 26.

[13] R. O. Duda, C. Avendano, and V. R. Algazi, "An Adaptable Ellipsoidal Head Model for the Interaural Time Difference," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 965–968 (Phoenix, AZ) (1999 Mar.). https://doi.org/10.1109/ICASSP.1999.759855.

[14] R. Bomhardt and J. Fels, "Analytical Interaural Time Difference Model for the Individualization of Arbitrary Head-Related Impulse Responses," presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), paper 9131.

[15] R. Bomhardt, M. Lins, and J. Fels, "Analytical Ellipsoidal Model of Interaural Time Differences for the Individualization of Head-Related Impulse Responses," *J. Audio Eng. Soc.*, vol. 64, no. 11, pp. 882–894 (2016 Nov.). https://doi.org/10.17743/jaes.2016.0041.

[16] H. Ziegelwanger and P. Majdak, "Modeling the Direction-Continuous Time-of-Arrival in Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 135, no. 3, pp. 1278–1293 (2014 Mar.). https://doi.org/10.1121/1.4863196.

[17] T. Cai, B. Rakerd, and W. M. Hartmann, "Computing Interaural Differences Through Finite Element Modeling of Idealized Human Heads," *J. Acoust. Soc. Am.*, vol. 138, no. 3, pp. 1549–1560 (2015 Sep.). https://doi.org/10.1121/1.4927491.

[18] N. Aaronson and W. M. Hartmann, "Testing, Correcting, and Extending the Woodworth Model for Interaural Time Difference," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 817–823 (2014 Feb.).

[19] X.-L. Zhong and B.-S. Xie, "A Novel Model for Interaural Time Difference Based on Spatial Fourier Analysis," *Chin. Phys. Lett.*, vol. 24, no. 5, pp. 1313–1316 (2007 May). https://doi.org/10.1088/0256-307X/24/5/052.

[20] C. P. Brown and R. O. Duda, "A Structural Model for Binaural Sound Synthesis," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 5, pp. 476–488 (1998 Sep.). https://doi.org/10.1109/89.709673.

[21] H. Gamper, M. R. P. Thomas, and I. J. Tashev, "Estimation of Multipath Propagation Delays and Interaural Time Differences From 3-D Head Scans," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 499–503 (South Brisbane, Australia) (2015 Apr.). https://doi.org/10.1109/ICASSP.2015.7178019.

[22] H. Gamper, M. R. P. Thomas, and I. J. Tashev, "Anthropometric Parameterisation of a Spherical Scatterer ITD Model With Arbitrary Ear Angles," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–5 (New Paltz, NY) (2015 Oct.). https://doi.org/10.1109/WASPAA.2015.7336941.

[23] H. Gamper, D. Johnston, and I. J. Tashev, "Interaural Time Delay Personalisation Using Incomplete Head Scans," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 461–465 (New Orleans, LA) (2017 Mar.). https://doi.org/10.1109/ICASSP.2017.7952198.

[24] M. Aussal, F. Alouges, and B. F. G. Katz, "ITD Interpolation and Personalization for Binaural Synthesis Using Spherical Harmonics," in *Proceedings of the AES UK 25th Conference: Spatial Audio in Today's 3D World* (2012 Mar.), paper 04.

[25] J. C. Middlebrooks, "Individual Differences in External-Ear Transfer Functions Reduced by Scaling in Frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1480–1492 (1999 Sep.). https://doi.org/10.1121/1.427176.

[26] AES, "AES Standard for File Exchange - Spatial Acoustic Data File Format," *AES Standard 69-2015* (2015 Jan.).

[27] A. Andreopoulou and B. F. G. Katz, "Identification of Perceptually Relevant Methods of Inter-Aural Time Difference Estimation," *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 588–598 (2017 Aug.). https://doi.org/10.1121/1.4996457.

[28] D. Schönstein and B. F. G. Katz, "HRTF Selection for Binaural Synthesis From a Database Using Morphological Parameters," in *Proceedings of 20th International Congress on Acoustics (ICA)*, pp. 1–6 (Sydney, Australia) (2010 Aug.).

[29] R. Pelzer, M. Dinakaran, F. Brinkmann, et al., "Head-Related Transfer Function Recommendation Based on Perceptual Similarities and Anthropometric Features," *J. Acoust. Soc. Am*, vol. 148, no. 6, pp. 3809–3817 (2020 Dec.). https://doi.org/10.1121/10.0002884.

[30] L. S. R. Simon, A. Andreopoulou, and B. F. G. Katz, "Investigation of Perceptual Interaural Time Difference Evaluation Protocols in a Binaural Context," *Acta Acust. united Acust.*, vol. 102, no. 1, pp. 129–140 (2016 Jan.). https://doi.org/10.3813/AAA.918930.

[31] Genelec, "Aural ID," https://www.genelec.com/aural-id (2021).

[32] A. Mäkivirta, M. Malinen, J. Johansson, et al., "Accuracy of Photogrammetric Extraction of the Head and Torso Shape for Personal Acoustic HRTF Modeling," presented at the *148th Convention of the Audio Engineering Society* (2020 May), paper 10323.

[33] S. Ise and M. Otani, "Real Time Calculation of the Head Related Transfer Function Based on the Boundary Element Method," in *Proceedings of the International Conference on Audio Display*, paper 69 (Kyoto, Japan) (2002 Jul.).

[34] M. Otani and S. Ise, "A Fast Calculation Method of the Head-Related Transfer Functions for Multiple Source Points Based on the Boundary Element Method," *Acoust. Sci. and Tech.*, vol. 24, no. 5, pp. 259–266 (2003 Sep.). https://doi.org/10.1250/ast.24.259.

[35] B. F. G. Katz, "Boundary Element Method Calculation of Individual Head-Related Transfer Function. I. Rigid Model Calculation," *J. Acoust. Soc. Am.*, vol. 110, no. 5, pp. 2440–2448 (2001 Nov.). https://doi.org/10.1121/1.1412440.

[36] B. F. G. Katz, "Boundary Element Method Calculation of Individual Head-Related Transfer Function. II. Impedance Effects and Comparisons to Real Measurements," *J. Acoust. Soc. Am.*, vol. 110, no. 5, pp. 2449–2455 (2001 Nov.). https://doi.org/10.1121/1.1412441.

[37] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," in *Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102 (New Paltz, NY) (2001 Oct.). https://doi.org/10.1109/ASPAA.2001.969552.

[38] P. Majdak, M. J. Goupell, and B. Laback, "3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training," *Atten. Percept. Psychophys.*, vol. 72, no. 2, pp. 454–469 (2010 Feb.). https://doi.org/10.3758/APP.72.2.454.

[39] R. Bomhardt, *Anthropometric Individualization of Head-Related Transfer Functions: Analysis and Modeling* (Logos Verlag Berlin GmbH, Berlin, Germany, 2017), 1st ed.

[40] B. F. G. Katz and M. Noisternig, "A Comparative Study of Interaural Time Delay Estimation Methods," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. 3530–3540 (2014 Jun.). https://doi.org/10.1121/1.4875714.

[41] J. Nam, J. S. Abel, and III J. O. Smith, "A Method for Estimating Interaural Time Difference for Binaural Synthesis," presented at the *125th Convention of the Audio Engineering Society* (2008 Oct.), paper 7612.

[42] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1.0: The Toolbox for Reproducible Research in Auditory Modeling," Acta Asustica, vol. 6, paper 19 (2022 May).

[43] H. Liu, Y. Fang, and Q. Huang, "Efficient Representation of Head-Related Transfer Functions With Combination of Spherical Harmonics and Spherical Wavelets," *IEEE Access*, vol. 7, no. 1, pp. 78214–78222 (2019 Jun.). https://doi.org/10.1109/ACCESS.2019.2921388.

[44] X. Zhong and B. Xie, "Head-Related Transfer Function and Virtual Auditory Display," in H. Glotin (Ed.), *Soundscape Semiotics: Localization and Categorization*, pp. 99–134 (IntechOpen, London, UK, 2014), 2nd ed. https://doi.org/10.5772/56907.

THE AUTHORS


Jaan Johansson


Aki Mäkivirta


Matti Malinen


Ville Saari

Jaan Johansson has been an R&D Specialist at Kuava Ltd., Kuopio, Finland since 2018. He received his Master of Science degree from the Department of Applied Physics at the University of Eastern Finland in 2019.

•

Aki Mäkivirta is an R&D Director at Genelec, Iisalmi, Finland. He received his Master of Science, Licentiate of Science, and Doctor of Science in Technology degrees in electrical engineering from Tampere University of Technology in 1985, 1989, and 1992, respectively. After research positions at the Medical Engineering Laboratory at Research Centre of Finland and Nokia Corporation Research Center, he joined Genelec Oy in 1995. Aki Mäkivirta is a Fellow of the AES and a life member of the Acoustical Society of Finland.

•

Matti Malinen is the managing director at Kuava Ltd., Kuopio, Finland. He received his Master of Science and Doctor of Philosophy degrees in applied physics from University of Kuopio in 2001 and 2004, respectively. After a brief post-doctoral researcher post at the University of Kuopio, he became a co-founder of Kuava Ltd. in 2007.

•

Ville Saari is an R&D Engineer at Genelec, Iisalmi, Finland. He received his Bachelor of Science and Master of Science degrees from Aalto University in 2011 and 2013, respectively. He also studied Music Technology at the National University of Ireland during the 2011–2012 semester. He joined Genelec in 2017 after a research engineer position at Fraunhofer IIS. Ville Saari is a Associate Member of AES.